



University  
of Glasgow

Pelle, R., Njahira, M.N., Osaso, J., Saya, R.M., Odongo, D.O., Toyé, P.G., Spooner, P.R., Musoke, A.J., Mwangi, D.M., Taracha, E.L.N., Morrison, W.I., Weir, W., Silva, J.C., and Bishop, R.P. (2011) *Two Theileria parva CD8 T cell antigen genes are more variable in buffalo than cattle parasites, but differ in pattern of sequence diversity*. PLoS ONE, 6 (4). e19015. ISSN 1932-6203

<http://eprints.gla.ac.uk/53485/>

Deposited on: 24 August 2012

# Two *Theileria parva* CD8 T Cell Antigen Genes Are More Variable in Buffalo than Cattle Parasites, but Differ in Pattern of Sequence Diversity

Roger Pelle<sup>1\*</sup>, Simon P. Graham<sup>1a</sup>, Moses N. Njahira<sup>1</sup>, Julius Osaso<sup>1b</sup>, Rosemary M. Saya<sup>1</sup>, David O. Odongo<sup>1</sup>, Philip G. Toye<sup>1</sup>, Paul R. Spooner<sup>1</sup>, Anthony J. Musoke<sup>1c</sup>, Duncan M. Mwangi<sup>1d</sup>, Evans L. N. Taracha<sup>1e</sup>, W. Ivan Morrison<sup>2</sup>, William Weir<sup>3</sup>, Joana C. Silva<sup>4</sup>, Richard P. Bishop<sup>1</sup>

**1** International Livestock Research Institute (ILRI), Nairobi, Kenya, **2** The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom, **3** Institute of Comparative Medicine, Glasgow University Veterinary School, Glasgow, United Kingdom, **4** Institute for Genome Sciences and Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, United States of America

## Abstract

**Background:** *Theileria parva* causes an acute fatal disease in cattle, but infections are asymptomatic in the African buffalo (*Syncerus caffer*). Cattle can be immunized against the parasite by infection and treatment, but immunity is partially strain specific. Available data indicate that CD8<sup>+</sup> T lymphocyte responses mediate protection and, recently, several parasite antigens recognised by CD8<sup>+</sup> T cells have been identified. This study set out to determine the nature and extent of polymorphism in two of these antigens, Tp1 and Tp2, which contain defined CD8<sup>+</sup> T-cell epitopes, and to analyse the sequences for evidence of selection.

**Methodology/Principal Findings:** Partial sequencing of the Tp1 gene and the full-length Tp2 gene from 82 *T. parva* isolates revealed extensive polymorphism in both antigens, including the epitope-containing regions. Single nucleotide polymorphisms were detected at 51 positions (~12%) in Tp1 and in 320 positions (~61%) in Tp2. Together with two short indels in Tp1, these resulted in 30 and 42 protein variants of Tp1 and Tp2, respectively. Although evidence of positive selection was found for multiple amino acid residues, there was no preferential involvement of T cell epitope residues. Overall, the extent of diversity was much greater in *T. parva* isolates originating from buffalo than in isolates known to be transmissible among cattle.

**Conclusions/Significance:** The results indicate that *T. parva* parasites maintained in cattle represent a subset of the overall *T. parva* population, which has become adapted for tick transmission between cattle. The absence of obvious enrichment for positively selected amino acid residues within defined epitopes indicates either that diversity is not predominantly driven by selection exerted by host T cells, or that such selection is not detectable by the methods employed due to unidentified epitopes elsewhere in the antigens. Further functional studies are required to address this latter point.

**Citation:** Pelle R, Graham SP, Njahira MN, Osaso J, Saya RM, et al. (2011) Two *Theileria parva* CD8 T Cell Antigen Genes Are More Variable in Buffalo than Cattle Parasites, but Differ in Pattern of Sequence Diversity. PLoS ONE 6(4): e19015. doi:10.1371/journal.pone.0019015

**Editor:** Gordon Langsley, Institut national de la santé et de la recherche médicale - Institut Cochin, France

**Received:** November 18, 2010; **Accepted:** March 25, 2011; **Published:** April 29, 2011

**Copyright:** © 2011 Pelle et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was funded with Grant No. R8042 from the Department for International Development, United Kingdom, and with International Livestock Research Institute core funds. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: r.pelle@cgiar.org

<sup>a</sup> Current address: Virology Department, Veterinary Laboratories Agency, Surrey, United Kingdom

<sup>b</sup> Current address: Africa Biosystems, Nairobi, Kenya

<sup>c</sup> Current address: ARC-OVI, Onderstepoort, South Africa

<sup>d</sup> Current address: Veterinary Medicine Research and Development, Pfizer, Kalamazoo, Michigan, United States of America

<sup>e</sup> Current address: Institute of Primate Research (IPR), Nairobi, Kenya

## Introduction

*Theileria parva* is a tick-transmitted protozoan parasite that causes East Coast fever (ECF), an economically important disease of cattle in eastern, central and southern Africa. The life cycle in the bovine host involves two intracellular stages: the schizont, which transforms lymphocytes and is responsible for disease pathology, and the intra-erythrocytic piroplasm, which is infective for the tick vector, *Rhipicephalus appendiculatus* [1]. The mammalian life cycle stages are haploid, but there is a transient diploid phase in the tick

gut after fusion of gametes [2] and sexual recombination has been observed between and within *T. parva* stocks during experimental transmission [3,4,5]. The African buffalo (*Syncerus caffer*) is the natural reservoir of *T. parva*, but the parasite does not cause disease in this species. Transmission of buffalo-derived *T. parva* to cattle results in a rapidly lethal disease, but in many cases the parasites do not differentiate to the erythrocyte-infective stage and are not transmissible by ticks [6,7]. Hence, although *T. parva* parasites that originate from buffalo are genotypically and antigenically closely related to *T. parva* maintained in cattle, available evidence

indicates that a significant proportion of the buffalo-derived population cannot be transmitted between cattle.

Immunisation of cattle by infection with *T. parva* and simultaneous treatment with long-acting tetracycline results in long term immunity against the homologous parasite strain but variable protection against challenge with heterologous parasite strains. Hence, vaccination of cattle in the field by this method requires a mixture of parasite strains. Studies of immune responses in cattle immunised by infection and treatment have provided evidence that MHC-I restricted CD8<sup>+</sup> T cells, which are able to kill parasitised lymphoblasts, are important mediators of immunity [8,9]. CD8<sup>+</sup> T-cell responses in cattle immunized with a single parasite isolate frequently exhibit parasite strain specificity [9,10] and in one study such strain specificity has been shown to correlate with immune status upon challenge with a heterologous strain [11]. These findings suggest that during parasite evolution there may have been immune-imposed selection for sequence diversity in the target antigens of the protective immune response.

Identification of molecular diversity in *T. parva* was initially based on application of a panel of monoclonal antibodies, mainly directed against a single immunodominant polymorphic antigen [12]. This was subsequently combined with detection of restriction fragment length polymorphisms in genomic DNA using probes derived from multi-copy gene families (reviewed in [4]). The recently sequenced *T. parva* genome has been used to identify a panel of DNA satellite markers allowing more comprehensive genotyping of parasite isolates [13]. Population studies of *T. parva* from different regions of Uganda and Kenya, using a subset of these markers, revealed a high level of diversity and a high frequency of infection of cattle with mixed genotypes [14,15]. Although some evidence of geographical sub-structuring was found among parasite populations, the lack of linkage disequilibrium between alleles at different loci was consistent with frequent genetic exchange.

Although these genotypic analyses have provided insight into the genetic structure of *T. parva* populations, they are uninformative with respect to the nature and selective pressures driving antigenic diversity relevant to immune protection. The recent identification of a number of *T. parva* antigens and epitopes recognised by CD8<sup>+</sup> T cells from *T. parva*-immune cattle [16,17] now provides an opportunity to address this question. Detailed studies of immune responses to two of these antigens, Tp1 and Tp2, have demonstrated that they are highly dominant targets of the CD8<sup>+</sup> T-cell response in cattle expressing the A18 and A10 class I MHC haplotypes, respectively. Preliminary analyses of Tp1 and Tp2 sequences in a small number of laboratory parasite strains revealed that the epitopes recognised by specific CD8<sup>+</sup> T cells are polymorphic and that this is associated with differential recognition by CD8<sup>+</sup> T cells [18].

The present study set out to determine the extent and nature of sequence diversity in genes encoding the Tp1 and Tp2 antigens in a series of isolates of *T. parva* obtained from cattle and buffalo in regions of Africa where ECF is endemic, and to interrogate the sequence data for evidence of the mode and strength of selection.

## Materials and Methods

### Ethics statement

The ILRI's Institutional Animal Care and Use Committee (IACUC) was established in 1993 to ensure that international standards for animal care and use are followed in all ILRI research involving use of animal. ILRI has complied voluntarily with the UK's Animals (Scientific Procedures) Act 1986 (<http://www.homeoffice.gov.uk/science-research/animal-research/>) that con-

tains guidelines and codes of practice for the housing and care of animals used in scientific procedures. The study reported here was carried out in strict accordance with the recommendations in the standard operating procedures of the ILRI IACUC and adequate consideration of the 3R's (Replacement of animal with non-animal techniques, Reduction in the number of animals used, and Refinement of techniques and procedures that reduce pain and distress). The ILRI Experimental Animal Request Form and Protocol for lymph node biopsies and blood collection was approved by the ILRI IACUC (IACUC ref no. 2006.9, IACUC ref 2006.10 and IACUC ref 2007.10).

### Parasite isolates

The 82 *T. parva* schizont-infected cell lines used in this study were established and maintained using previously described methodologies [19]. These comprised lines derived from four different sources (Table S1).

**(i) Laboratory samples (LS).** Nine cell lines were generated by infection of bovine lymphocytes *in vitro* with sporozoite stabilates of established laboratory isolates [19,20] while eight were established from biopsies from animals infected with these stabilates (Table S1-A). All except two of these parasites were originally derived from cattle in Kenya, Uganda, Zimbabwe and Zambia: the two exceptions were from buffalo, one from Ol Pejeta ranch, Nanyuki, Kenya and the second from an animal experimentally infected with a parasite originating from a buffalo in Serengeti National Park Tanzania [21,22]. Stabilate 4108 was prepared following tick passage from the Marikebuni *T. parva* stabilate 3014 [21]. Three cell lines (LS6, 7 and 8), BN64 Muguga, BN65 Kiambu 5 and BN140 Serengeti [23,24] were generated by infecting lymphocytes with sporozoite seed stabilates used in the production of a large batch of the Muguga trivalent cocktail employed for vaccination by infection and treatment [25].

**(ii) Cattle-derived (CD).** A further 27 cell lines (Table S1-B) were isolated directly from clinically reacting infected cattle in three regions of Kenya between 1997 and 2000. Kilifi in coast province; Kakuzi in central province [26] and Nyairo in the Trans-Nzoia district in western Kenya [15]. Two cell lines cloned by limiting dilution (CD4 and CD5) were derived from a mixed parasite isolate from Kakuzi [26].

**(iii) Buffalo-derived (BD).** A third set of 16 isolates was obtained directly from buffalo (Table S1-C). Ten isolates (BD1 to 9 and BD13) were obtained by culture of leukocytes from blood samples collected from buffalo in the Masai Mara game reserve and Trans-Mara, Kenya [27,28]. Isolate D10 was obtained from a buffalo infected by application of ticks from Ngong near Nairobi. Isolate BD11 was from buffalo 5641 born in captivity at the Central Veterinary Laboratories (CVL, Nairobi) and not exposed to *Theileria* prior to infection with *T. p. lawrencei* stabilate 177 [27]. BD12 and 14 represented parasites originating from Laikipia district in central Kenya while BD13 and 15 were from Mara, and all held at the CVL in Nairobi. For number BD16, PBMC from clean buffalo 7344 was infected with stabilate 3081 prepared from buffalo 7014, followed by cloning as described [21,29].

**(iv) Buffalo-associated (BA).** The fourth set of 22 isolates was obtained from cattle that grazed in close proximity to buffalo at Marula farm in Naivasha, Kenya (Table S1-D). The cattle were part of a trial performed in the year 2000 to investigate protection afforded to immunized cattle in an area where cattle and buffalo were co-grazing. The cattle were immunized by infection and treatment (ITM) with either the FA01 Muguga cocktail stabilate, or one of two different Marikebuni stabilates 316 and 3014 (a parent stabilate of 316). Un-immunized control cattle were also exposed to *T. parva* challenge. Following vaccination, cattle were

allowed to graze in close proximity to buffalo. They therefore received a *T. parva* challenge from ticks that may have recently fed on buffalo. In the study reported here, cell lines were obtained from lymph node biopsies from 22 of these cattle during the acute reaction phase, which exhibited the typical clinical pattern of low parasitosis and parasitaemia associated with infection by parasites originating from buffalo [1]. The majority of these parasites contained a 129 bp sequence insert within the p67 sporozoite antigen gene, assayed by PCR (Bishop R, Spooner P, Musoke AJ, Odongo D, unpublished), which is typical of *T. parva* parasites of buffalo origin in East Africa [30].

### CD8 epitopes present within the selected antigen genes

We have previously identified two antigens 543 and 174 amino acids long, Tp1 and Tp2 respectively, that are recognised by bovine CD8<sup>+</sup> T cells (GenBank accession nos. [XP\\_762973](#) and [XP\\_765583](#)). This was achieved by direct screening of a *T. parva* Muguga schizont cDNA library with CD8 T-cell lines. A single CD8<sup>+</sup> T-cell epitope (VGYPKVKEEML: Tp1<sub>214–224</sub>) has been mapped in the Tp1 antigen [16]. Six distinct CD8 T-cell epitopes, SHEELKKGML- Tp2<sub>27–37</sub>, DGFDRDALF- Tp2<sub>40–48</sub>, KSSHG MGKVGK- Tp2<sub>49–59</sub>, FAQSLVCVL- Tp2<sub>96–104</sub>, QSLVCVL MK- Tp2<sub>98–106</sub> and KTSIPNPCKW- Tp2<sub>138–147</sub>, presented by 5 different class I alleles have been mapped in Tp2 [17,31,32].

### Antigen cloning and sequencing strategy

Specific forward and reverse primers were designed for PCR amplification of Tp1 and Tp2 gene sequences; one primer pair amplified a 432 bp region located between nucleotides 523 and 954 of the Tp1 ORF containing the known CD8 T-cell epitope (Tp1 forward primer: 5'-ATGGCCACTTCAATTGCATTTGCC-3'; Tp1 reverse primer: 5'-TTAAATGAAATATTTATGAGCTTC-3') containing a tagged stop codon underlined; a second primer pair was used to amplify 525 bases comprising the complete Tp2 ORF (Tp2 forward primer: 5'-ATGAAATTGGCCGCCAGATTA-3'; Tp2 reverse primer: 5'-CTATGAAGTGCCGGAGGCTTC-3'). Total genomic DNA from *T. parva*-infected lymphocytes was prepared as recommended [33]. Genomic DNA (30 ng) was PCR amplified in a 33 µl reaction with 25 U/ml AmpliTaq Gold DNA polymerase (Applied Biosystems, USA) in the presence of 33 ng each of specific forward and reverse primers, 1× PCR Gold buffer (Applied Biosystems) containing 1.5 mM MgCl<sub>2</sub> and 200 µM dNTP and using a programmable thermal cycler (MJ Research, Watertown, MA, USA). The cycling conditions were: step 1, 95°C for 11 min; step 2, 95°C for 30 s; step 3, 50°C for 45 s; step 4, 72°C for 30 s (30 times from steps 2 to 4); step 5, 72°C for 10 min. 7 µl of the PCR products was analyzed by electrophoresis in ethidium bromide-stained 1.5% agarose gels as described [33]. All the samples studied were positively amplified by PCR. For sequencing, 10 µl of the PCR products was treated with 10 U of exonuclease I and 1 U of shrimp alkaline phosphatase (United States Biochemical, USA) at 37°C/15 min, followed by 15 min incubation at 80°C. 4 µl of treated PCR product was sequenced directly or following cloning into pGEM-T Easy vector (Promega, USA) using specific primers and an ABI 3730 capillary sequencer (Applied Biosystems).

### Predicted amino acid sequences

Open reading frames present within the sequences generated from the amplified DNA fragments were translated into amino acid sequences using EMBOSS-Transeq software [34] and converted into FASTA format. Alignments of nucleotide and

amino acid sequences were performed using CLUSTALW version 1.83 [35].

### Genetic diversity and population structure

The evolutionary genetic distances (expressed in terms of the number of differences per 100 bases or amino acids, including length polymorphisms) between every pair of sequences in a multiple alignment were generated using the DISTMAT program accessible at <http://hpc.ilri.cgiar.org/emboss/> [36]. Estimates of DNA polymorphism,  $\pi$ , determined as the average number of nucleotide differences per site, were obtained with DnaSP v5 [37]. The Excel plug-in 'Genalex6' [38] was used to perform Principal Component Analysis (PCA) based on these distances. Analysis of molecular variance (AMOVA) was performed using 'Genalex6' [38] in order to investigate the distribution of genetic variation among allelic sequences and to determine the level of population differentiation. Pair-wise estimates of genetic distance among populations within the species were calculated using  $\Phi_{PT}$ , the proportion of variance among populations relative to total variance.

### Molecular evolution

The unrooted phylogenetic tree for each locus was estimated by neighbor-joining as implemented in MEGA [39]. The mode and intensity of selection,  $\omega$  ( $\sim d_N/d_S$ , the ratio of non-synonymous to synonymous substitution rates), acting on each locus was estimated using a codon-based substitution model [40] as implemented in PAML's program *codeml* [41]. Several models of evolution, namely M0, M1a, M2, M3, M7 and M8 [42] were tested to determine which provided a better fit to the data, given the phylogenetic relationships among sequences. Model M0 assumes a single rate of evolution for all codons (sites); all other models implemented allow for variable  $\omega$  values across sites; however they do not allow  $\omega$  to vary among branches of the phylogeny. For each model analyses were run twice, with different  $\omega$  start values, to account for the possibility of sub-optimal peaks in the likelihood function. In each case, the analysis with the highest likelihood score is reported. A likelihood ratio test (LRT) was used to determine the significance of the difference in likelihood value of pairs of nested models, in which one allows for positive selection and the other does not (namely, M2 *vs.* M1a, and M8 *vs.* M7). In each case the test statistic (two times the difference in likelihood of two models of evolution  $x$  and  $y$ ,  $2\Delta\ln L_{M_y-M_x}$ ) was evaluated against a chi-square distribution with 2 degrees of freedom. All analyses were launched and monitored, and results visualized, using IDEA [43].

### Results

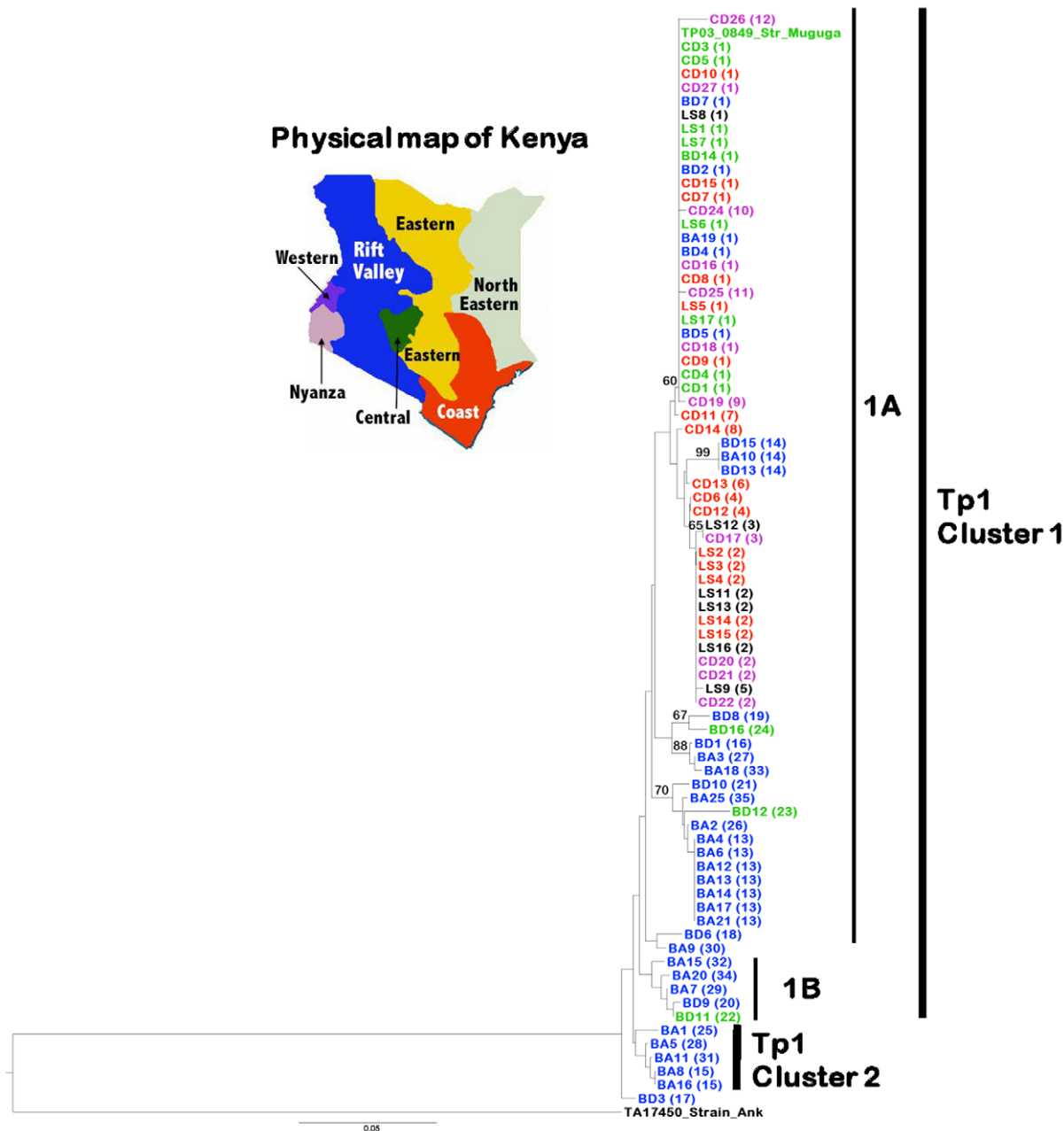
In order to determine the level of genetic diversity in isolates of *T. parva* obtained from cattle and buffalo in regions of Africa where ECF is endemic, sequence polymorphism analyses of two CD8 T cell target antigens, Tp1 and Tp2, were conducted. Identifiers prefixed with LS or CD denote cattle parasites with no association with buffalo, while BD or BA denote those from buffalo or cattle with association with buffalo (see Materials and Methods for details).

#### The Tp1 locus

The 432 bp region of Tp1 that was sequenced is located in the centre of the gene and extends from nucleotides 523 to 954 of the reference Tp1 sequence (accession number [XP\\_762973](#)), corresponding to sample LS1 (Table S1). This region encodes 144 amino acids or 26.5% of the 543-residue Tp1 protein. We identified 35 different alleles of the Tp1 gene in the 79 parasites

whose sequence was determined (Tables S1 and S2); overall polymorphism in the region was  $\pi = 0.019$  (or 1.9%). We were unable to determine the Tp1 sequence from 3 of the 82 isolates. The alleles were distinguished by single nucleotide polymorphisms (SNPs) at 51 nucleotides, and two in-frame indels of 36 and 12 nucleotides, respectively (Figure S1). Compared to the reference sequence, deletions were present in 22 isolates, among which there were three buffalo-derived parasites (BD3, 8 and 16) in which both deletions were observed. Allele 1, which is present in the *T. parva* Muguga reference sequence (LS1), was represented

in 24 of the 79 isolates (30.3%). The alleles that were genetically most distant from the Muguga sequence were alleles 17, 19 and 24, found in BD isolates. These shared a genetic distance of 24.82% from allele 1. However, this was heavily influenced by the presence of a deletion of 36 nucleotides in these isolates (Figure S1). The largest observed genetic distance of 26.76% was between BD isolates 10 and 16 that were derived from buffalo in south-western Kenya (Mara) and central Kenya (Laikipia), respectively (Fig. 1). In comparison, the greatest genetic distance observed between any pair of isolates derived from cattle that had



**Figure 1. Neighbour-joining tree showing relationships among 79 cattle and buffalo-derived *T. parva* isolates.** These data are based on the sequences of 35 Tp1 gene alleles obtained in this study. *T. annulata* (TA17450 Strain\_Ank) was used as an outgroup to root the tree. Isolates with their corresponding alleles in parentheses are indicated and colour-coded based on their geographical location in Kenya. Isolates from outside Kenya (Tanzania, Uganda, Zambia and Zimbabwe) are depicted in black. The TP03\_0849 gene from the *T. parva* (Muguga) genome sequence was also included in the analysis. Bootstrap values  $>50\%$  are shown above branches. A Major clade and minor clade with two sub-clusters is indicated. doi:10.1371/journal.pone.0019015.g001



no contact with buffalo was only 2.3% (isolates CD26 versus CD17, LS9 and LS12).

We observed that four different Tp1 alleles (1, 2, 3 and 5) were present among the 16 LS parasites (derived originally from nine stocks from five countries), 11 different alleles among the 25 CD stocks from Kenya, 11 among the 16 Kenyan BD stocks, and 15 among the 25 BA stocks that came from a single farm (Table S1; Fig. 1). A total of 12 different alleles was obtained from the 41 LS and CD parasites, and overall polymorphism in these isolates was  $\pi = 0.7\%$ . In contrast, there were 24 different alleles from the 41 BA and BD stocks studied, with  $\pi = 2.2\%$ . This suggests that there is greater diversity within the Tp1 locus among parasites in buffalo than in those maintained in cattle.

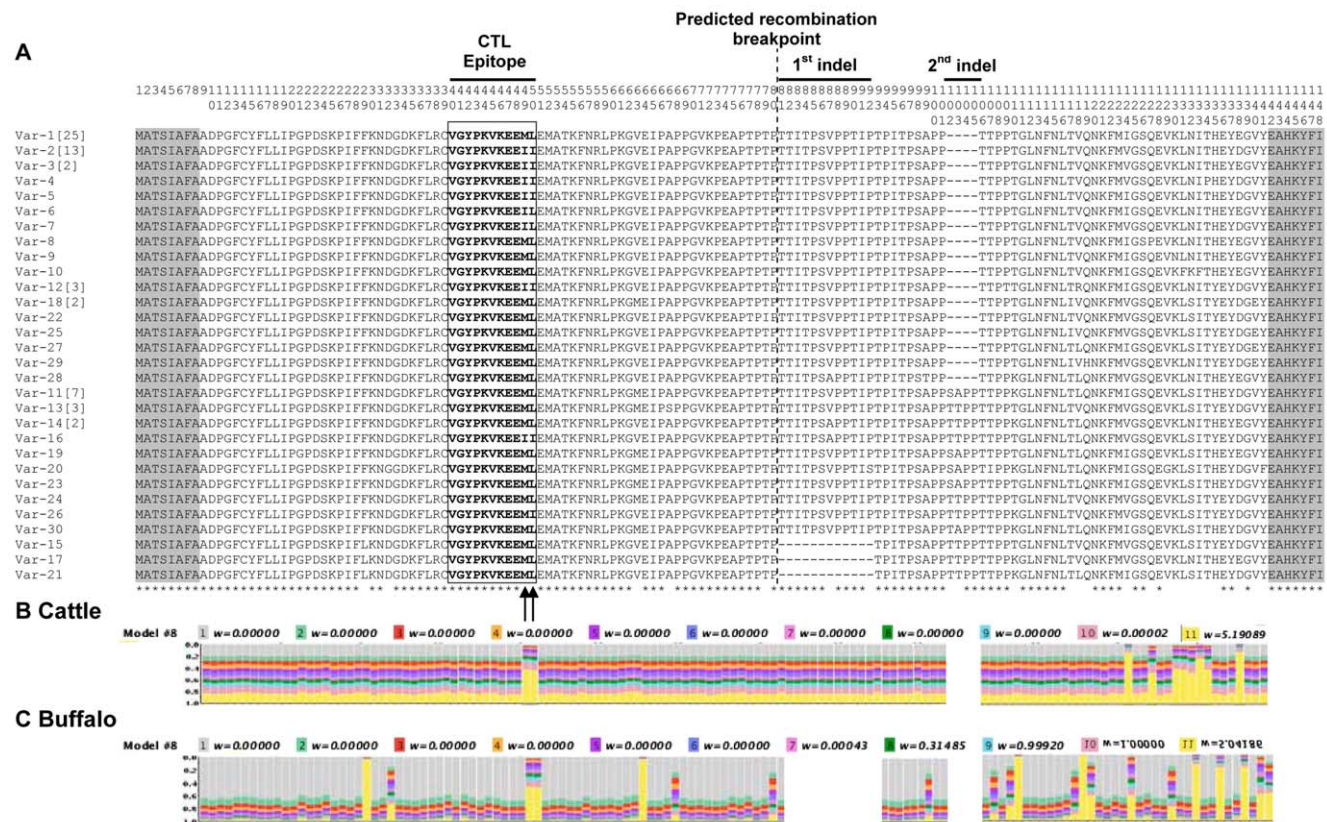
The predicted protein sequences of the Tp1 alleles were compared and the nucleotide variations described above resulted in 30 distinct protein variants (Table S2), due to amino acid changes at 27 residues, among which 4 were located in the indel regions (Fig. 2). The breadth of variation at the protein level reflected that observed for the alleles, in that greater diversity was present in the BA and BD isolates. Only one allele (variant 1 which was present in the Muguga reference sequence) was found in all four groups (Table S1).

Of particular interest was a comparison of the defined CD8<sup>+</sup> T-cell epitope (VGYPKVK E EML) located within the sequenced

region of Tp1. We observed three variants of the epitope resulting from substitutions in the two carboxy-terminal amino acids (Fig. 2). The three new variant epitopes ended in -MI, -IL or -II, as indicated in Table S1. The majority of isolates analysed (58 out of 79) displayed the epitope sequence (-ML) present in the *T. parva* Muguga reference isolate, including 36 of the 41 BA and BD isolates. The next most common variant (-II) was observed in 21 cell lines, and was strongly associated with isolates derived from cattle (LS and CD). The single substitution variants (-IL and -MI) were observed in two CD and one BA cell line, respectively.

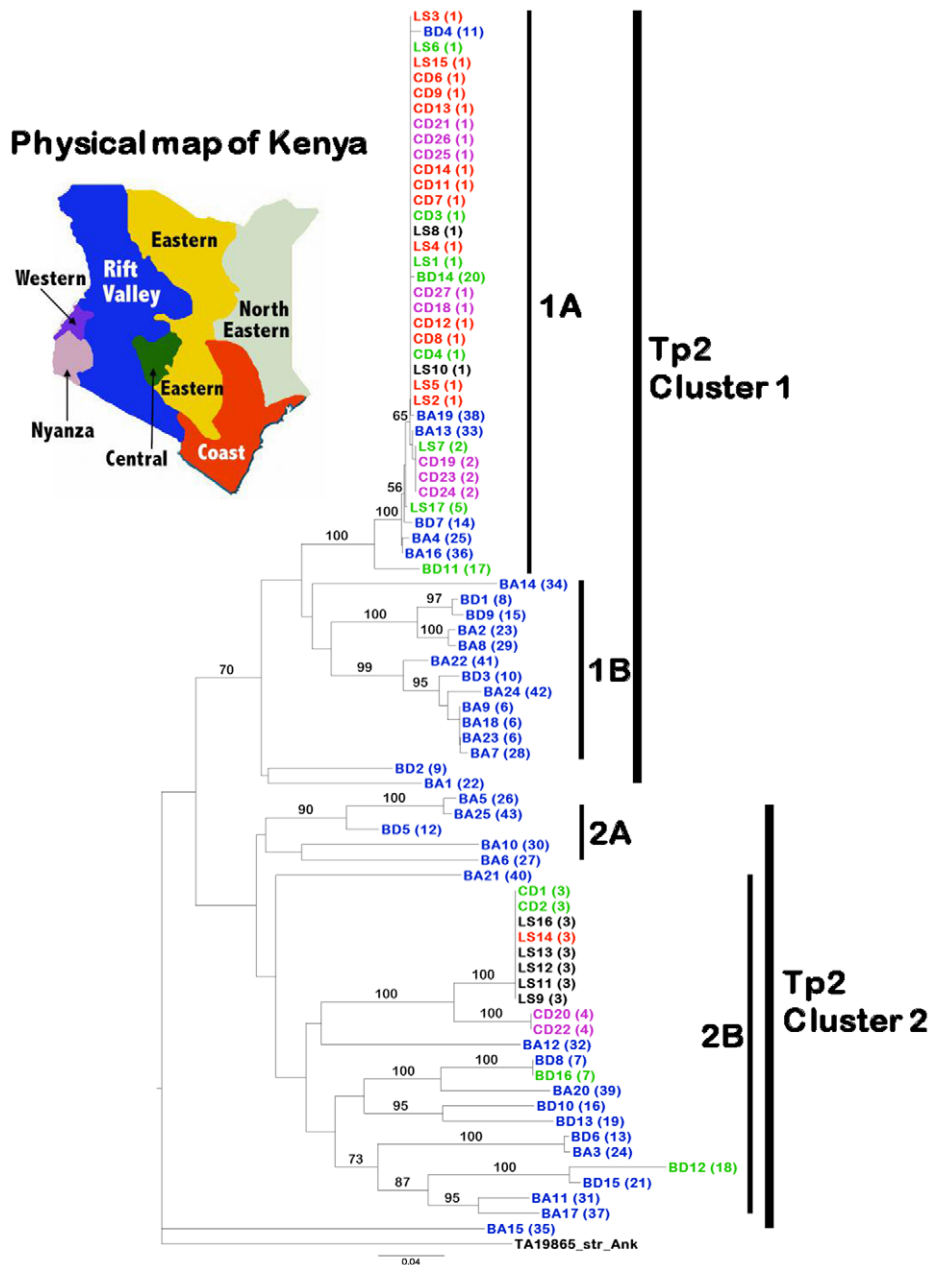
## The Tp2 locus

The full-length Tp2 gene sequence in the reference Muguga genome sequence (XP\_765583), predicted to encode a protein of 174 amino acids, was sequenced in 77 isolates. We were unable to determine the sequence from five isolates. As four Tp2 sequences were identified in the Marula buffalo-associated isolate N106 (BA21 to 24), a total of 80 Tp2 sequences were analyzed (Table S1). We identified 43 alleles among the sequenced isolates, with SNPs observed at 320 nucleotides (Fig. 3 and Figure S2), and overall  $\pi = 17.9\%$ , almost an order of magnitude higher than in the Tp1 segment analyzed. No indels were detected. Among the cattle-derived isolates, four different alleles were present among



**Figure 2. Multiple amino acid sequence alignment of the 30 Tp1 antigen variants obtained in this study.** (A) The single letter amino acid code is used throughout. Variants named Var1-30. Residue coordinates are listed above the alignment. The single currently identified CD8 T-cell target epitope (coordinates 40–50 bolded and boxed) and the positions of the two indels are indicated. The two polymorphic residues in the T-cell epitope are indicated by arrows. A predicted recombination breakpoint between amino acid residues 80 and 81 is indicated with a broken vertical line. The frequency of each variant is indicated in square brackets, when larger than 1. Residues conserved in all sequences are identified below the alignment (\*). The flanked PCR primers regions not included in estimations of % residues conserved are shadowed. The distribution of selective constraints across the Tp1-encoded peptide, according to model M8, is shown for cattle (B) and buffalo (C) parasites. The stacked bar chart above each amino acid residue shows the probability that the degree of selective constraint on that residue falls into each of the several  $\omega$  classes indicated by the color legend.

doi:10.1371/journal.pone.0019015.g002



**Figure 3. Neighbour-joining tree showing relationships among 80 cattle and buffalo-derived *T. parva* isolates.** These data are based on the sequences of 43 Tp2 alleles obtained in this study. The Tp2 homologous from *T. annulata* (TA19865\_strain\_Ank) was used as an outgroup to root the tree. Isolates with their corresponding alleles in parentheses are indicated and colour-coded based on their geographical location in Kenya. Isolates from outside Kenya (Tanzania, Uganda, Zambia and Zimbabwe) are shown in black. Bootstrap values >50% are shown above branches. Two major clades and sub-clusters within these are indicated. doi:10.1371/journal.pone.0019015.g003

the 17 LS isolates. Three of these alleles, plus a fifth were observed among the 22 CD isolates. In contrast, 15 different alleles were detected among the 16 BD parasites and 23 different alleles were identified among the 25 BA isolates. Thus among 39 isolates from cattle that were not associated with buffalo only 5 Tp2 alleles were observed (with allele 1 from the Muguga reference sequence the most common, being present in 24 of the 39 isolates (61.5%)), while among 41 isolates or clones of likely buffalo origin there were 38 alleles (Table S1). The BA and BD parasites shared no alleles with the LS and CD stocks. The Muguga allele was not present in any of the BD or BA parasites, although the allele from BA19

differs by a single, synonymous nucleotide substitution. The sequence of BD12 was genetically the most distant from the Muguga reference sequence, the genetic distance being 29.71%. The largest genetic distance of 31.24% was observed between BA15 and two other buffalo related isolates, BA17 and BD13. In contrast, isolates BD8 and 16 had identical Tp2 alleles (Fig. 3; Table S1). This was also the case for BA9, 18 and 23 although they carried different Tp1 alleles. The greatest genetic distance between any pair of cattle-derived isolates was 24.95%, which is 10-fold greater than the maximum genetic distance observed between any pair of Tp1 sequences from cattle.







**Table 1.** Tp2 CTL epitopes variants obtained in this study.

Epitope Variants	Epitope 1 (Tp2 <sub>27–37</sub> ) (25 variants)	Epitope 2 (Tp2 <sub>40–48</sub> ) (21 variants)	Epitope 3 (Tp2 <sub>49–59</sub> ) (23 variants)	Epitope 4 (Tp2 <sub>96–104</sub> ) (19 variants)	Epitope 5 (Tp2 <sub>98–106</sub> ) (19 variants)	Epitope 6 (Tp2 <sub>138–147</sub> ) (18 variants)
V1	SHEELKKGML 1, 5, 11, 14	DGFDRLALF 1, 2, 5, 11, 14, 17, 22, 24, 28, 32, 35	KSSHGMGKVGK 1, 2, 5, 14, 17, 24, 28, 32, 35	FAQLVCVL 1, 2, 5, 11, 14, 17, 24, 32, 35	QSLVCVLMK 1, 2, 5, 11, 14, 24, 32, 35	KTSIPNPCKW 1, 2, 5, 11, 14, 22, 32, 35, 38
V2	SDEELNKLGM 2, 34	PDLKKNRLF 3, 4	LTSHGMRIGR 3	FAASIKCVA 3	ASIKCVAQY 3	KPSVNPNCW 3, 4
V3	SDDELDTLGM 3	DGFEKERLF 6, 8, 10, 15, 27, 39, 40	LTSHGMRIGR 4	LAASIKCVS 4	ASIKCVSHH 4	KEDVPNPCKW 6, 8, 10, 15, 22, 27, 28, 40
V4	SDNELDTLGL 4	PDPVKERLF 7	KTSHSMGMIGK 6, 10, 27, 39, 40	FGQSVVCVL 6, 10, 27, 39, 40	QSVVCVLMR 6, 10, 27, 39, 40	KQSILNPCTW 7, 17
V5	SDEELKKGML 6, 10, 17, 24, 27, 32, 35, 39	EGLDKDALF 9	LTSKAMTTVGK 7	FAQSIKCVS 7, 37	QSIKCVSQH 7, 37	KTNIPNPCKW 9
V6	SDEEELSLGM 7	EGFDKEALF 12	RSSHGMGKVGK 8, 15	FVESILCVI 8, 15	ESILCVIKK 8, 15	KSNIPNPCKW 12
V7	SDEELKKGML 8, 15	PGFDKEVLF 13, 23	KSSKGMGKVGK 9	FAQSIKCVS 9	QSIKCVAQH 9, 25, 41	NNNINLPCKW 13, 23
V8	SNEELKKGML 9, 12	PDPDKQRLF 16	KSSHGMGEVGK 11	FGQSIKCVS 12, 25, 41	QSIKCVQK 12	KGDVPNPCKW 16, 19
V9	SDDELKKGML 13, 23	SDPDRETLF 18	KSSKGMTKVGK 12	FGQSIKCVS 13, 23, 30, 36	QSIKCVQK 13, 23, 30, 36	TSDVPNPCEW 18, 20
V10	SHSELETLGM 16	PDPDKETLF 19, 20	KSSKAMTTTGK 13, 23	YAQSIKCVS 16	QSIKCVANN 16, 19	KGDVPNPCKW 21, 25, 41
V11	TPPELEALGRL 18	EGFDREALF 21	LTSKAMSTVGK 16	FGASIQCVV 18, 20	QSLVCVLMN 17	KDNTNPNCW 26
V12	SHEELKKGML 19	EGLDMEALF 25, 41	KTSKAMTMTGR 18, 19, 20	CFAQSIKCVS 19	ASIQCVVKN 18, 20	KGDAPNPCTW 29
V13	TEELKKGML 20	DGFDKELLF 26	KTSNGMTKVGK 21	FGQSLKCVL 21	QSLKCVLQH 21	KEDVPNPCEW 30, 36
V14	TSEELKKGML 21	DGFDQRRLF 29	KSSHGMGKVGK 22	FVESIMCVI 22, 28	ESIMCVIKK 22, 28	KPKIPNPCKW 31
V15	SDDELKKGML 22, 28	DNLDKDKLF 30	KTSKGMTKVGK 25, 41	FAQSIKCVV 26, 29	QSIKCVVKN 26, 29	TSDIPNPCKW 33
V16	TNEELKKGML 25, 41	PHPDKERLF 31	KSSKGMTAVGK 26, 38	FAASIKCVS 31	ASIKCVSNK 31	KDKIPNPCKW 34
V17	NDELKKGML 26	DGFDRELLF 33	LTSKSMSEVGR 29	LAQSIKCVV 33	QSIKCVVSK 33	KQSILNPNCW 37
V18	SDEELKKGML 29	PDFEKEKLF 34	MTSKAMTATGR 30	FAQSLKCVL 34	QSLKCVLQH 34	KEDVPNPCKW 39
V19	TEDELKKGML 30	QDPDKETLF 36	KTSKGMTKVGK 31	FAQSIKCVS 38	KSIKCVSQH 38	
V20	NDEELENLGM 31	PDPNKERLF 37	KSSHGMGKIGR 33			
V21	TVEELREMGML 33	PGFDKELLF 38	AASHGLGKVGK 34			
V22	SDDELNKLGM 36		KTSKAMTATGR 36			
V23	TDEELENLGM 37		LTSKSMMTVGK 37			
V24	TEDELKNMGLI 38					
V25	SDELKKGML 40					

Note: Numbers following the epitope sequences are the corresponding antigen alleles (see Table S1) carrying the epitopes. Up to 8 different amino acid residues (in bold and underlined) are used in the second amino acid position in epitope 6.  
doi:10.1371/journal.pone.0019015.t001

## Molecular evolution analysis of the Tp1 and Tp2 genes

In order to characterize the evolutionary forces acting on these two loci, we estimated  $\omega$  ( $\sim d_N/d_S$ ), the ratio of non-synonymous to synonymous substitution rates, which provides a measure of the mode and intensity of selection. For each locus the isolates were divided into two groups, one containing parasites obtained from cattle (CD and LS), and the other all parasites derived from, or associated with, buffalo (BD and BA, respectively). Six models of evolution (M0, M1a, M2, M3, M7 and M8) were tested (see Methods), to determine which among them provides a better fit to the patterns of variation observed across each gene. While M0 is clearly an oversimplification for most proteins, it provides a good benchmark against which more complex models can be compared. Three of the models tested (M2, M3 and M8) allow a subset of residues to evolve under positive selection. The comparison of two of those models, M2 and M8, with models that assume the absence of positive selection (M1a and M7, respectively) using a likelihood ratio test (LRT) provides a direct assessment of the likelihood of positive selection (see Methods).

**(i) The Tp1 gene.** The evolution of the Tp1 locus from both cattle and buffalo parasites is best described by models that allow positively selected residues, namely M2, M3 and M8, as indicated by higher log likelihood values relative to the other models (Table 2). Model M3 has the highest likelihood score for both cattle and buffalo parasites. This model allows one subset of residues to evolve under positive selection with  $\omega$  estimated from the data ( $\omega_2$ ), and the remaining residues grouped into two classes 0 and 1, each with  $\omega \leq 1$  estimated from the data ( $\omega_0$  and  $\omega_1$ ). In the case of cattle parasites, 80% of residues fall into classes 0 ( $p_0 = 0.77$ ) or 1 ( $p_1 = 0.03$ ), both of which evolve under extreme purifying selection ( $\omega_0 \approx \omega_1 \approx 0$ ) (Table 2). The remaining 20% of residues ( $p_2 = 0.20$ ) appear to evolve under strong positive selection, with non-synonymous polymorphism occurring about five times more frequently than synonymous polymorphisms ( $\omega_2 \sim 5.2$ ). Ten of these residue positions show statistically significant evidence for positive selection, and two of them (sites with coordinates 49 and 50 in the protein alignment) fall within the known Tp1 epitope (Table 2, Fig. 1b). Models M2 and M8 have nearly identical likelihoods to

**Table 2.** Evolution of the Tp1 locus: likelihood value (lnL) and parameter estimates for six evolution models implemented.

Model <sup>a</sup>	lnL	$d_N/d_S$ <sup>b</sup>	Estimated parameter values <sup>c</sup>
<b>CD+LS isolates</b>			
M0 (one-ratio)	-620.2986	0.91	$\omega = 0.91$
M1a (nearly neutral)	-619.0294	0.55	$p_0 = 0.45, (p_1 = 0.55), \omega_0 = 0.00$
M2 (positive sel.)	-617.4723	1.07	$p_0 = 0.79, p_1 = 0.0, (p_2 = 0.21)$ $\omega_0 = 0.00, \omega_2 = 5.19$ Positively selected sites (BEB): 124, 133, 138
M3 (discrete)	-617.4722	1.07	$p_0 = 0.77, p_1 = 0.03, (p_2 = 0.20)$ $\omega_0 = 0.00, \omega_1 = 0.00, \omega_2 = 5.19$ Positively selected codons (NEB): 49**, 50**, 124**, 127**, 130**, 131**, 132**, 133**, 134**, 138**
M7 (beta)	-619.0489	0.50	$p = 0.005, q = 0.005$
M8 (beta & $\omega$ )	-617.4725	1.07	$p_0 = 0.79, (p_1 = 0.21), p = 0.005, q = 1.702$ $\omega = 5.19$ Positively selected sites (BEB): 49, 50, 124, 127, 130, 131, 132, 133, 134, 138
<b>BA+BD isolates</b>			
M0 (one-ratio)	-892.7489	0.49	$\omega = 0.49$
M1a (nearly neutral)	-876.3387	0.29	$p_0 = 0.71, (p_1 = 0.29), \omega_0 = 0.00$
M2 (positive sel.)	-870.9605	0.59	$p_0 = 0.72, p_1 = 0.21, (p_2 = 0.07)$ $\omega_0 = 0.016, (\omega_1 = 1.0), \omega_2 = 5.13$ Positively selected sites (BEB): 29, 63, 110, 118**, 132, 135, 138
M3 (discrete)	-870.9524	0.58	$p_0 = 0.68, p_1 = 0.25, (p_2 = 0.07)$ $\omega_0 = 0.00, \omega_1 = 0.84, \omega_2 = 4.98$ Positively selected codons (NEB): 29, 63, 110, 118**, 132, 135, 138
M7 (beta)	-876.3474	0.30	$p = 0.005, q = 0.012$
M8 (beta & $\omega$ )	-870.9584	0.58	$p_0 = 0.93, (p_1 = 0.07), p = 0.021, q = 0.066$ $\omega = 5.04$ Positively selected sites (BEB): 29*, 49, 50, 63*, 110*, 118**, 132, 135, 138

<sup>a</sup>Sites models of evolution. Nomenclature as in Yang *et al.* (2000). Model with highest likelihood for each locus is shown in *italics*.

<sup>b</sup>Weighted average of  $d_N/d_S$  across all sites.

<sup>c</sup>Parameters in parenthesis are inferred (probabilities add up to 1); all others are estimated from the data. Significance levels indicate probability that  $\omega > 1$  for the site:

\* $p \geq 0.95$ ;

\*\* $p \geq 0.99$  (no asterisks signifies  $p < 95\%$ ). Significance levels for M3 is based on a Naïve Empirical Bayes (NEB) analysis, while that of models M2 and M8 is based in a Bayes Empirical Bayes (BEB) analysis (Yang, Wong, and Nielsen 2005).

doi:10.1371/journal.pone.0019015.t002

model M3 (all have  $\ln L \sim -617.5$ ). Model M2, similarly to model M3, groups residues into three classes, but one of those classes, class 1, is composed of residues assumed to be strictly conserved ( $\omega = 0$ ). Model M8 is an approximation to a continuous distribution, with residues partitioned into 10 classes, each with  $\omega < 1$ , and another class with  $\omega \geq 1$ . Averaged across the whole segment of the gene, and according to the three most likely models (M2, M3 and M8) the number of synonymous and non-synonymous polymorphisms per site is roughly identical ( $\omega = 1.07$ ), suggesting either the absence of selective constraints ( $\omega = 1$ ) or positive selection ( $\omega > 1$ ). To distinguish between these scenarios we used a likelihood ratio test of nested models. When M2 is compared to M1a, and M8 to M7, the models M2 and M8, which allow positive selection, do not provide a significantly better fit to the data than M1a and M7 ( $2\Delta\ln L_{M2-M1a} = 3.11$ ,  $P \sim 0.21$ ;  $2\Delta\ln L_{M8-M7} = 3.15$ ,  $P \sim 0.21$ ), suggesting that the accumulation of non-synonymous mutations may be due to the absence of selective constraints. As discussed above, the cattle sequences are very similar to each other; in fact, they are clustered in only two clades, each with a very recent common ancestor (Fig. 1).

In Tp1 sequences from buffalo parasites, and again according to model M3, 68% of the sites ( $p_0$ ) evolve under very strong purifying selection ( $\omega = 0$ ), while  $\sim 25\%$  of the sites have almost identical numbers of synonymous and non-synonymous polymorphisms ( $\omega_1 = 0.84$ ). Finally, about 7% of the sites evolve under strong purifying selection, with  $\omega_2 \sim 5$ . Seven of those sites show statistically significant positive selection, but none of them falls within the epitope region. Across the whole segment analyzed, according to model M3,  $\omega \sim 0.6$ . Again, models M2 and M8 have nearly identical values of log likelihood ( $\ln L \sim -870.0$ ) and  $d_N/d_S$  ( $\omega = 0.6$ ) to model M3. However, in contrast to cattle parasites, in the case of the sequences determined from buffalo parasites the models allowing positive selection (M2 and M8) seem to provide a significantly better fit to the data than their nested models (M1 and M7, respectively), with  $2\Delta\ln L_{M2-M1a} = 10.76$  ( $P \sim 0.046$ ) and  $2\Delta\ln L_{M8-M7} = 10.78$  ( $P \sim 0.0046$ ). This result suggests that positive selection has been a significant force in the evolution of this locus in buffalo parasites.

Overall, the results for Tp1 reveal a remarkably high proportion of non-synonymous polymorphisms, with  $d_N/d_S > 0.5$  in cattle and

buffalo parasites, respectively. Interestingly, however, those polymorphisms are concentrated in a minority of sites, with strong purifying selection prevalent in 70% to 80% of the residues in buffalo and cattle parasites, respectively (Table 2, Fig. 1bB and 1C).

**(ii) The Tp2 gene.** Similarly to the observations for Tp1, the results for locus Tp2 from both cattle and buffalo parasites show that models that allow positively selected sites (namely, M2, M3 and M8) provide a better fit to the data than do models M0, M1a and M7, as implied by their log likelihood values (Table 3). Model M3 provides the best fit to the cattle parasite data, but M2 and M8 have just slightly lower log likelihood values. According to model M3, approximately 24% of the sites ( $p_1+p_2$ ) evolve under positive selection, but for none of those sites is the probability of  $\omega>1$  larger than 95% (Table 3). Model M8 provides the best description for the evolution of buffalo parasite sequences, but with a likelihood value clearly higher than all other five models, including M2 and M3 (Table 3). According to this model, ~12% of the sites fall within the class of positively selected residues, and the result is statistically significant for residue positions 46, 77, 91, 130 and 139 (Table 3, Fig. 4C).

Comparisons of the two nested pairs of models, M2 vs M1a and M8 vs. M7, show statistically strong evidence of positive selection in parasite sequences from cattle ( $2\Delta\ln L_{M2-M1a} = 9.03$ ,  $P \sim 0.011$ ;  $2\Delta\ln L_{M8-M7} = 10.50$ ,  $P \sim 0.005$ ) and from buffalo ( $2\Delta\ln L_{M2-M1a} = 29.21$ ,  $P \sim 4.5 \times 10^{-7}$ ;  $2\Delta\ln L_{M8-M7} = 28.38$ ,  $P \sim 6.9 \times 10^{-7}$ ). Unlike for Tp1 locus, the comparison between nested models is significant in cattle. Finally, the average  $\omega$  is high in both cattle ( $0.78 < \omega < 1.29$ ) and buffalo ( $0.57 < \omega < 0.68$ ), as determined by the three most likely evolution models (Table 3).

In sharp contrast to what is observed for Tp1 (Fig. 2B–C), in which >50% of amino acid residues are extremely conserved ( $\omega = 0$ ), the Tp2 locus does not have a class of residues with an estimated  $\omega = 0$  (Fig. 4B–C) in either buffalo or cattle parasite sequences. Even though the average  $\omega$  is about the same in both loci, polymorphism is an order of magnitude higher in Tp2 ( $\pi = 17.9\%$ , compared with  $\pi = 1.9\%$  in Tp1), and the observed variation widely distributed across the Tp2 locus, with the majority of sites either evolving relatively rapidly ( $\omega > 0.5$ ) or under positive selection ( $\omega > 1$ ). In spite of this difference, for both loci the average  $\omega$  is larger than 0.5. Together with the evidence for positive selection in several residues, and the known antigenic role of the

**Table 3.** Evolution of the Tp2 locus: likelihood values (lnL) and parameter estimates for the evolutionary models implemented.

Model <sup>a</sup>	lnL	$d_N/d_S^b$	Estimated parameter values <sup>c</sup>
<b>CD+LS isolates</b>			
M0 (one-ratio)	−1213.7784	0.45	$\omega = 0.45$
M1a (nearly neutral)	−1205.9230	0.45	$p_0 = 0.66$ , ( $p_1 = 0.34$ ), $\omega_0 = 0.17$
M2 (positive sel.)	−1201.4097	0.86	$p_0 = 0.65$ , $p_1 = 0.33$ , ( $p_2 = 0.02$ ) $\omega_0 = 0.20$ , $\omega_2 = 19.41$ Positively selected sites (BEB): 83, 105, 126, 130, 165*
M3 (discrete)	−1201.3067	1.29	$p_0 = 0.76$ , $p_1 = 0.22$ , ( $p_2 = 0.02$ ) $\omega_0 = 0.25$ , $\omega_1 = 1.41$ ; $\omega_2 = 50.26$ Positively selected sites (NEB): 29, 34, 39, 40, 46, 49, 66, 69, 74, 77, 81, 83, 86, 91, 98, 101, 104, 105, 106, 122, 126, 130, 165
M7 (beta)	−1206.8250	0.47	$p = 0.4024$ , $q = 0.4527$
M8 (beta & $\omega$ )	−1201.5758	0.78	$p_0 = 0.97$ , ( $p_1 = 0.03$ ), $p = 0.695$ , $q = 0.865$ $\omega = 12.24$ Positively selected sites (BEB): 83, 105, 126, 130, 164**
<b>BA+BD isolates</b>			
M0 (one-ratio)	−6194.3698	0.50	$\omega = 0.50$
M1a (nearly neutral)	−5989.1350	0.54	$p_0 = 0.54$ , ( $p_1 = 0.46$ ), $\omega_0 = 0.14$
M2 (positive sel.)	−5974.5287	0.68	$p_0 = 0.50$ , $p_1 = 0.41$ , ( $p_2 = 0.09$ ) $\omega_0 = 0.14$ , ( $\omega_1 = 1.0$ ), $\omega_2 = 2.24$ Positively selected sites (BEB): 46, 73, 77, 83, 91*, 118, 130*, 139**, 146, 149, 165
M3 (discrete)	−5967.0899	0.57	$p_0 = 0.41$ , $p_1 = 0.40$ , ( $p_2 = 0.19$ ) $\omega_0 = 0.09$ , $\omega_1 = 0.60$ , $\omega_2 = 1.55$ Positively selected sites (NEB): 20, 34, 38*, 39, 40, 42, 46**, 56, 65, 68, 72, 73**, 74, 77**, 78, 82, 83*, 86, 91**, 94, 101*, 104, 105, 118*, 130**, 139**, 146**, 148, 149**, 165**
M7 (beta)	−5976.5388	0.46	$p = 0.4114$ , $q = 0.4877$
M8 (beta & $\omega$ )	−5962.3507	0.57	$p_0 = 0.88$ , ( $p_1 = 0.12$ ), $p = 0.487$ , $q = 0.711$ $\omega = 1.76$ Positively selected sites (BEB): 20, 34, 38, 39, 42, 46*, 56, 65, 69, 73, 74, 77*, 78, 82, 83, 86, 91*, 94, 101, 104, 118, 130*, 139**, 146, 149, 165

**Note:** Legend as in Table 2.

doi:10.1371/journal.pone.0019015.t003

proteins encoded by these loci, these results suggest that positive selection has contributed significantly for the retention of non-synonymous polymorphisms in the parasite populations. Finally, it is of interest to note that, in both loci, the average  $\omega$  value is slightly higher among cattle than in buffalo parasites.

### Distribution of *T. parva* antigen variability among parasite populations

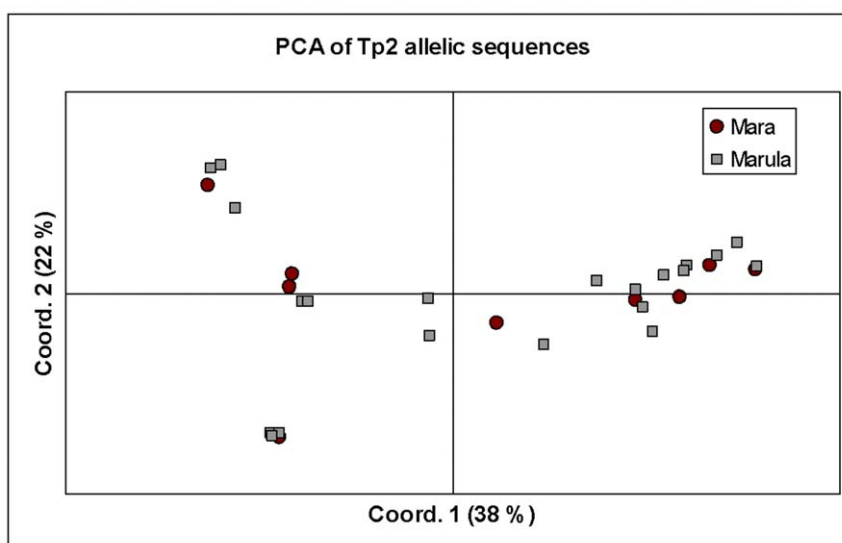
We generated neighbour-joining trees for both loci rooted with the orthologous sequences from *T. annulata*, to examine whether the sequence diversity observed in Tp1 and Tp2 was associated with geographical origin or mammalian host species. In the case of Tp1, the phylogenetic analysis separated the 35 distinct alleles into a small clade (Tp1 cluster 2) containing 5 sequences originating exclusively from buffalo-associated parasites and a major clade (Tp1 cluster 1) containing two sub-clades 1A and 1B (Fig. 1). An exception was isolate BD3 which represented the single unique allele 17 that did not group with any of these clades or sub-clades. Allele 1 found in the reference Muguga F100 isolate and 67 other isolates, among which were the three component stocks of the FAO1 Muguga cocktail live vaccine (LS6, 7 and 8), was grouped with 23 other alleles into the subcluster 1A. This group contained 68 out of 79 isolates (86%) of both cattle and buffalo origins and from different geographical regions of Kenya and as well as from Tanzania (LS8), Uganda (LS12, 16), Zambia (LS9, 10) and Zimbabwe (LS11, 13).

The phylogenetic analysis of the 43 Tp2 distinct alleles also separated the 80 isolates studied into two major clusters (Fig. 3). Despite relatively large overall polymorphism among cattle isolates ( $\pi = 10.7\%$ ) most differences are attributed to variations between sequences in cluster 1 and cluster 2 (Fig. 5). Within each cluster the cattle parasites are nearly identical, and hence only 5 different alleles were found. In contrast, overall polymorphism among buffalo parasites was  $\pi = 21.2\%$ . Most of the alleles were within cluster 1, which grouped into two subclusters 1A (37 isolates from LS, CD, BD and BA groups) and 1B (12 isolates, all derived from buffalo). The three isolates BD2, BA1 and BA15 encoding single unique alleles 9, 22 and 35 respectively, did not fall into any clade.

Allele 1 found in the reference Muguga F100 isolate LS1 as well as in 23 other isolates was grouped with ten other alleles into the subcluster 1A. As was observed with the Tp1 subcluster 1A, the Tp2 alleles in cluster 1A were represented in parasites from different ecological regions throughout Kenya, including the coast (Mariakani, Marikébuni and Kilifi, with the isolates LS2-5; CD6-9, 11-14), the central region (Muguga, Kakuzi, Kiambu, Laikipia and Nanyuki, with the isolates LS1, 6, 7 and 17; CD3 and 4; BD11 and 14) and the western region/Rift valley (Nyairo and Masai Mara with the isolates CD18, 25-27, BD4). In addition, Tp2 alleles from an isolate from Zambia (LS10) and from the three component stocks of the FAO1 Muguga cocktail live vaccine (containing isolates from Kenya and Tanzania) were grouped into the 1A subcluster. Cluster 2 alleles comprised two subclusters, 2A (5 isolates, all derived from buffalo) and 2B (23 isolates from the LS, CD, BD and BA groups). The alleles in subcluster 2B were from parasites isolated from Kenya, Uganda, Zambia and Zimbabwe. The presence, for both loci, of large clades of nearly identical sequences from cattle isolates suggests a rapid spread among cattle.

We further analysed the partition of genetic diversity in the Tp2 locus, using AMOVA, by comparing the alleles from the Mara buffalo-derived ( $n = 10$ ) parasites (BD) with those from the parasites isolated from buffalo-associated cattle (BA) ( $n = 25$ ). All of the variation among Tp2 alleles in the locus was found to exist within each population, with none of the variation in the dataset attributable to differences between the populations ( $\Phi_{PT} = 0$ ). Thus, buffalo-derived isolates from Mara showed no evidence of differentiation from isolates from cattle at Marula farm that were presumed to have been infected by ticks that had fed on buffalo. The relationship between buffalo-derived (BD) and buffalo-associated (BA) isolates is depicted using a PCA plot (Fig. 5).

Three remarkable observations result from these analyses. First, the laboratory (LS) and cattle-derived (CD) isolates represent a very small proportion of the variation in each locus, even though the samples were collected from across Kenya and surrounding countries. Secondly, for both Tp1 and Tp2, LS and CD isolates are grouped into two clades nested within more ancient lineages



**Figure 5. Principal component analysis of Tp2 allelic diversity.** This diagram illustrates the relationship between buffalo-derived (Mara) and buffalo-associated, cattle-derived isolates (Marula), suggesting the isolates may belong to a single population. The proportion of variation in the dataset explained by the 1<sup>st</sup> and 2<sup>nd</sup> principal components is indicated in parenthesis.  
doi:10.1371/journal.pone.0019015.g005



comprising the buffalo-derived and buffalo-associated parasites (Fig. 1 and 4c). Even though all buffalo-associated parasites (BA) were collected on a single farm within a short period of time, they displayed extensive sequence diversity in Tp1 and Tp2, suggesting that the vast majority of antigen variability resides in parasites maintained in buffalo. These data are consistent with the hypothesis that *T. parva* maintained in cattle represents a subset of the overall *T. parva* population present in buffalo and that buffalo act as a reservoir of antigenically diverse parasites that can be transmitted to cattle sharing grazing grounds with buffalo. Finally, despite the fact that the parasites examined were derived from two different mammalian hosts and from geographically distant sites, predominantly in Kenya but with a few isolates from other countries, the phylogenetic analyses suggest that they represent a single parasite population. More detailed population genetic studies using neutral loci are required to confirm this hypothesis.

## Discussion

The study revealed extensive allelic diversity in the Tp1 and Tp2 genes resulting in polymorphisms in all seven CD8 T-cell epitopes identified within these antigens. Overall polymorphism was one order of magnitude higher in Tp2 than in Tp1, and at least twice as high in buffalo than in cattle parasites. As expected, at the protein level the Tp2 antigen showed more extensive sequence diversity than Tp1 and both antigens displayed greater diversity in parasite isolates from buffalo or cattle grazing with buffalo compared to parasites from cattle in buffalo-free areas. While analyses of the sequences demonstrated evidence of selection for amino acid changes including residues within T cell epitopes, there was no significant enrichment for positively selected residues in the epitope sequences.

Nucleotide sequence analysis of part of the Tp1-encoding gene revealed 35 distinct alleles defined by point mutations and two regions containing indels. The predicted protein sequences identified 30 different variants of the antigen. Only one of these (variant 1, which is found in the reference *T. parva* Muguga isolate) was shared by the cattle-derived and buffalo-derived isolates. The single CD8<sup>+</sup> T-cell epitope identified in Tp1 was relatively conserved at the amino acid level when compared to the six known Tp2 epitopes, with only four variants observed among 79 isolates, involving coding changes at residues 10 and 11 of the epitope.

Sequence comparisons indicated that the Tp2 antigen is highly polymorphic, with 43 distinct alleles and 41 predicted protein variants observed among the 80 Tp2 sequences obtained. The sequence variation was distributed across the gene and no deletions or insertions were observed. Analysis of the sequences encoding the Tp2 CD8<sup>+</sup> T-cell epitopes revealed a very high number of amino acid substitutions, with only two or three amino acid residues being conserved in each of the six epitopes across the 80 *T. parva* isolates/clones studied. Eight different amino acids were found at a single residue position of one epitope (Table 1).

As might be expected from the difference in epitope number between the two loci, the frequency of variable residues was much higher in Tp2 than in Tp1. However, although there was evidence that positively selected codons were present in the T-cell epitopes, the epitope regions were not enriched for such codons compared to the rest of the antigen. Given that these antigens enter the class I MHC processing pathway and that cattle representing a relatively limited percentage of bovine haplotype space have so far been evaluated for the immunodominant *T. parva* epitopes that they present, it is likely that additional epitopes, recognised in the context of different MHC backgrounds are yet to be discovered. If

the antigens contain a large number of additional epitopes, the analyses as conducted herein might fail to reveal significant enrichment of selection within the epitope sequences. Alternatively, these antigens may not be predominant targets of CD8<sup>+</sup> T-cell responses in the buffalo, in which much of parasite evolution is likely to have occurred. A third explanation is that the host CD8<sup>+</sup> T-cell response is not the major factor responsible for selection of diversification in these two proteins in *T. parva*. The Tp1 and Tp2 proteins have no identifiable orthologues in parasites other than *Theileria*, or any other taxa. Their functions remain unknown. Hence, these genes may be subject to alternative selective pressures, acting either on the schizont or other life cycle stages, such as those present in the tick vector, which result in the observed sequence diversity. If the above hypothesis were correct, escape of parasites from T cell recognition, in the mammalian host, would therefore be fortuitous, as has been proposed for both *Plasmodium* [44] and *Theileria* [45].

The sequence diversity data obtained in this study has provided further insight into the population structure of *T. parva*. The Tp2 alleles in parasites isolated from cattle that shared grazing with buffalo at a single farm in the Rift Valley of Kenya showed much greater sequence diversity than those from cattle where buffalo were not present, although these were derived from geographically distant sites. They were also indistinguishable from the alleles in the sample of buffalo-derived parasites. This strongly supports the assertion that parasites isolated from cattle grazing with buffalo originated from the buffalo reservoir. The observation that the diversity of Tp2 is considerably higher in buffalo ( $\pi = 21\%$ ) than in cattle not in contact with buffalo ( $\pi = 11\%$ ) suggests that variation in Tp2 evolved and is maintained primarily in the wildlife reservoir. The observed diversity may represent polymorphisms that accumulated in the wildlife (buffalo/tick) reservoirs over several millennia. This would have occurred well before the introduction of cattle into the region, estimated from archaeological records to be approximately six thousand years ago [46]. The limited diversity in the cattle population can be explained by a founder effect, in which only a subset of the buffalo parasite gene pool can be established and transmitted within the cattle population. Although the parasites isolated from cattle associated with buffalo are highly diverse, as illustrated by the level of polymorphism in the Tp2 gene in BA isolates, these parasites were isolated from lymph node biopsies obtained from clinically reacting animals. The cattle died with a typical buffalo-derived *T. parva* clinical syndrome involving low schizont parasitaemia and a low piroplasm parasitaemia. This clinical picture (combined with the presence of insertions in the p67 sporozoite antigen gene characteristic of buffalo parasites [30]) suggests that the Marula BA parasites represent a sample of cattle-infective parasites present in the local buffalo population, rather than parasites that can sustainably be transmitted within the cattle population by ticks. Although it is not possible to make conclusive statements regarding the slightly higher average  $\omega$  in cattle vs. buffalo parasites given the current sample size, it is interesting to speculate that it might be the result of more recently imposed selection associated with adaptation to a new host in evolutionary terms, relative to the buffalo. Alternatively, this difference may reflect the rapid expansion of the *T. parva* population among cattle, with a concomitant relaxation of purifying selection.

There is a striking contrast in the number of Tp1 and Tp2 variants observed among parasites isolated from cattle relative to buffalo. Of the 30 amino acid sequence variants identified in the partial sequence of Tp1, 10 are present in cattle not associated with buffalo, and nine of these are cattle-specific. By contrast only four out of 41 antigenic variants in Tp2 are present in cattle

isolates and one of these was obtained only from a cattle maintained parasite (LS17) that was originally isolated from a buffalo. There are several possible explanations for this observation, including that subsequent to transfer from buffalo there has been (1) a relaxation of purifying selection in Tp1, (2) an increase in diversifying selection in Tp1 in cattle populations, or (3) an increase in purifying selection in the Tp2 locus. Finally, it is also possible that the founder population exhibited more diversity in Tp1 than Tp2, (4) either by chance or (5) because the allelic variants compatible with maintenance in cattle are more limited in Tp2 than in Tp1. The latter hypothesis (5), as well as (3), might imply that adaptation to cattle is influenced by the Tp2 allele sequence. The accumulation of both synonymous and non-synonymous mutations in cattle parasites sampled may reflect mostly an episode of relaxed selection associated with a rapid expansion of these isolates in the cattle population. Alternatively, the recent common ancestry of all cattle isolates may have resulted in lack of analytical power, due to insufficient time for new variants to arise and for the effects of small selection coefficients associated with slightly deleterious or advantageous mutations to be felt.

An interesting aspect of both Tp1 and Tp2 diversity is that the data suggest an epidemic population structure in cattle that are not co-grazing with buffalo, where a limited number of genotypes appears to have expanded and become over-represented in the population. For both loci, sequences from cattle parasites are clustered in only two clades, each composed of closely related 1 alleles. This population model has been observed previously in regional populations in Uganda, based on data from a panel of variable number tandem repeat (VNTR) polymorphic markers [14]. A more recent study demonstrated significant linkage disequilibrium within three regional populations in Kenya [15]. In both studies there was evidence of extensive genetic diversity among isolates but in Kenya there was no direct evidence of clonal expansion of particular genotypes. However, the analyses of the Kenyan populations were based on data from a large panel of VNTR polymorphic markers and did not include antigen gene sequencing; therefore the results are not directly comparable to those presented herein.

Allelic polymorphism of target antigens in natural populations of *T. parva* is a key issue for understanding the basis of strain specificity in immunity to the parasite. This study has focused on comparative analyses of the sequences of two genes that encode dominant CD8<sup>+</sup> T-cell target antigens in field isolates of *T. parva* from cattle and buffalo. The principal aims were to provide information on antigenic diversity at the population level and to investigate whether the antigens are subject to positive selection for amino acid changes. The demonstration of more extensive sequence diversity in buffalo-derived parasites than in those maintained in cattle is consistent with the idea that a limited subset of the *T. parva* population has become adapted for maintenance by tick transmission in cattle, or that many parasite genotypes cannot be maintained in the cattle tick transmission cycle. These findings are also consistent with results of vaccination studies, which have provided evidence that immunity induced by infection and

treatment with a mixture of parasite isolates is not always effective against challenge with buffalo-derived parasites [47]. Although there was clear evidence of positive selection for amino acid changes in both antigens, these were not significantly enriched within the known CD8<sup>+</sup> T-cell epitope regions. Additional functional studies of the CD8<sup>+</sup> T-cell epitopes, currently underway, will identify those alleles that are able to escape CD8<sup>+</sup> T cell recognition and, by identifying the amino acid substitutions that are critical for escape, may shed further light on whether these antigens are subject to selection by cellular immunity.

## Supporting Information

**Figure S1 Multiple sequence alignment of 35 Tp1 alleles obtained in this study.** The single CD8 T-cell epitope is overlined (plain line). The two polymorphic nucleotides in the epitope domain are shadowed. Positions of flanked residues in the Tp1 gene fragment are numbered. There are 3 size-polymorphic Tp1 of 444, 432 and 408 nucleotides, respectively. The two indels are overlined with a broken line (deletion) and a dotted line (insertion). (\*) indicates identical residues. The frequency of each allele is indicated in square brackets, when larger than 1. The flanked PCR primers regions are boxed.

(PDF)

**Figure S2 Multiple sequence alignment of 43 Tp2 alleles obtained in this study.** The frequency of each allele is indicated in square brackets, when larger than 1. Positions of flanked residues in the Tp2 gene fragment are numbered. (\*) indicates identical residues. The flanked PCR primers regions are boxed.

(PDF)

**Table S1 Cell lines infected with *T. parva* cattle-derived parasite stocks isolated from different geographic areas of eastern and southern Africa. A. Laboratory samples. B. Kenyan field isolates from cattle with no association with buffalo. C. Isolates derived from buffalo.**

(DOC)

**Table S2 Tp1 gene alleles and their corresponding antigen variants.**

(DOC)

## Acknowledgments

We acknowledge the staff involved in ILRI animal sampling, the ITM vaccine and tissue culture groups, large animal and tick units and biological services, for excellent animal care and provision of parasites.

This is ILRI publication no. IL-201005.

## Author Contributions

Conceived and designed the experiments: RP RPB. Performed the experiments: RP SPG MNN JO. Analyzed the data: RP WW JCS RPB. Contributed reagents/materials/analysis tools: RP RMS DOO PRS AJM DMM ELNT. Wrote the paper: RP PGT WIM WW JCS RPB.

## References

- Norval RAI, Perry BD, Young AS (1992) The epidemiology of *Theileria* in Africa. London: Academic.
- Gauer MU, Mackenstedt H, Mehlhorn E, Schein F, Zapf E, et al. (1995) DNA measurements and ploidy determination of developmental stages in the life cycles of *Theileria annulata* and *T. parva*. Parasitol Res 81: 565–574.
- Morzaria SP, Young JR, Spooner PR, Dolan TT, Bishop RP (1993) *Theileria parva*: a restriction map and genetic recombination. In: Morzaria SP, ed. Genome analysis of protozoan parasites. ILRAD: Nairobi. pp 67–73.
- Bishop R, Geysen PD, Skilton R, Odongo D, Nene V, et al. (2002) Genomic polymorphism sexual recombination and molecular epidemiology of *Theileria parva*. In: McKeever D, Dobbelaere D, eds. Kluwer: Academic Press. pp 32–40.
- Katzer F, Ngugi D, Oura C, Bishop RP, Taracha EL, et al. (2006) Extensive genotypic diversity in a recombining population of the apicomplexan parasite *Theileria parva*. Infect Immun 74: 5456–5464.
- Morrison WI, Goddeeris BM, Brown WC, Baldwin CL, Teale AJ (1989) *Theileria parva* in cattle: characterization of infected lymphocytes and the immune responses they provoke. Vet Immunol Immunopathol 20: 213–237.
- Uilenberg G (1999) Immunization against diseases caused by *Theileria parva*: a review. Trop Med Int Health 4: A12–20.

8. McKeever DJ, Taracha ELN, Innes EL, MacHugh ND, Awino E, et al. (1994) Adoptive transfer of immunity to *Theileria parva* in the CD8<sup>+</sup> fraction of responding efferent lymph. *Proc Natl Acad Sci USA* 91: 1959–1963.
9. Taracha ELN, Goddeeris BM, Morzaria SP, Morrison WI (1995) Parasite strain specificity of precursor cytotoxic T cells in individual animals correlates with cross-protection in cattle challenged with *Theileria parva*. *Infect Immun* 63: 1258–1262.
10. Goddeeris BM, Morrison WI, Toye PG, Bishop R (1990) Strain specificity of bovine *Theileria parva* specific cytotoxic T cells is determined by the phenotype of the restricting class I MHC. *Parasitology* 69: 38–44.
11. Taracha EL, Goddeeris BM, Morzaria SP, Morrison WI (1995) Parasite strain specificity of precursor cytotoxic T cells in individual animals correlates with cross-protection in cattle challenged with *Theileria parva*. *Infect Immun* 63: 1258–1262.
12. Toye PG, Metzelaar MJ, Wijngaard PL, Nene V, Iams K (1995) Characterization of the gene encoding the polymorphic immunodominant molecule a neutralizing antigen of *Theileria parva*. *J Immunol* 155: 1370–1381.
13. Oura CA, Odongo DO, Lubega GW, Spooner PR, Tait A, et al. (2003) A panel of microsatellite and minisatellite markers for the characterisation of field isolates of *Theileria parva*. *Int J Parasitol* 33: 1641–1653.
14. Oura CA, Asiimwe BB, Weir W, Lubega GW, Tait A (2005) Population genetic analysis and sub-structuring of *Theileria parva* in Uganda. *Mol Biochem Parasitol* 140: 229–239.
15. Odongo DO, Oura CA, Spooner PR, Kiara H, Mburu D, et al. (2006) Linkage disequilibrium between alleles at highly polymorphic mini- and micro-satellite loci of *Theileria parva* isolated from cattle in three regions of Kenya. *Int J Parasitol* 36: 937–946.
16. Graham SP, Pellé R, Honda Y, Mwangi DM, Tonukari NJ, et al. (2006) *Theileria parva* candidate vaccine antigens recognized by immune bovine cytotoxic T lymphocytes. *Proc Natl Acad Sci USA* 103: 3286–3291.
17. Graham SP, Pellé R, Yamage M, Mwangi DM, Honda Y, et al. (2008) Characterization of the fine specificity of bovine CD8 T-cell responses to defined antigens from the protozoan parasite *Theileria parva*. *Infect Immun* 76: 685–694.
18. MacHugh ND, Connelly T, Graham SP, Pelle R, Formisano P, et al. (2009) CD8<sup>+</sup> T cell responses to *Theileria parva* are preferentially directed to a single dominant antigen: Implications for parasite strain-specific immunity. *Eur J Immunol* 39: 2459–2469.
19. Brown CGD (1979) Propagation of *Theileria*. In: Maramorosch K, Hirumi H, eds. *Practical Tissue Culture Applications* Academic Press. pp 223–245.
20. Goddeeris BM, Morrison WI (1988) Techniques for the generation cloning and characterization of bovine cytotoxic T cells specific for the protozoan *Theileria parva*. *J Tissue Cult Methods* 11: 101–110.
21. Morzaria SP, Dolan TT, Norval RAI, Bishop RP, Spooner PR (1995) Generation and characterisation of cloned *Theileria parva* parasites. *Parasitology* 111: 39–49.
22. Geysen D, Bishop R, Skilton R, Dolan TT, Morzaria S (1999) Molecular Epidemiology of *Theileria parva* in the field. *Trop Med Int Health* 4: A21–27.
23. Radley DE, Brown CGD, Burridge MJ, Cunningham MP, Kimber CD, et al. (1975b) East Coast fever. 3 Chemoprophylactic immunization of cattle using oxytetracycline and a combination of theilerial strains. *Vet Parasitol* 1: 51–60.
24. Young AS, Branagan D, Brown CGD, Burridge MJ, Cunningham MP, et al. (1973) Preliminary observations on a theilerial species pathogenic to cattle isolated from buffalo (*Syncerus caffer*) in Tanzania. *Br Vet J* 129: 382–389.
25. Morzaria S, Spooner P, Bishop R, Mwaura S (1999) The preparation of a composite stabilate for immunisation against East Coast fever. In: Morzaria S, Williamson S, eds. *Live vaccines for Theileria parva: Deployment in Eastern Central and Southern Africa* International Livestock Research Institute: Kenya. pp 56–61.
26. Musoke A, Rowlands J, Nene V, Nyanjui J, Katende J, et al. (2005) Subunit vaccines based on the p67 major surface protein of *Theileria parva* sporozoites reduce severity of infection derived from field tick challenge. *Vaccine* 23: 3084–3095.
27. Conrad PA, Stagg DA, Grootenhuys JG, Irvin AD, Newson J, et al. (1987) Isolation of parasites from African Buffalo (*Syncerus caffer*) and characterization with anti-schizont monoclonal antibodies. *Parasitology* 94: 413–423.
28. Baldwin CL, Malu MN, Grootenhuys JG (1988) Evaluation of cytotoxic lymphocytes and their parasite strain specificity from African buffalo infected with *Theileria parva*. *Parasite Immunology* 10: 393–403.
29. Grootenhuys JG, Letich BL, Stagg DA, Dolan TT, Young AS (1987) Experimental induction of *Theileria parva* lawrencei carrier state in an African buffalo (*Syncerus caffer*). *Parasitology* 94: 425–431.
30. Nene V, Musoke A, Gobright E, Morzaria S (1996) Conservation of the sporozoite p67 vaccine antigen in cattle-derived *Theileria parva* stocks with different cross-immunity profiles. *Infect Immun* 64: 2056–2061.
31. Akoolo L, Pellé R, Saya R, Awino E, Nyanjui J, et al. (2008) Evaluation of the recognition of *Theileria parva* vaccine candidate antigens by cytotoxic T lymphocytes from Zebu cattle. *Vet Immunol Immunopathol* 12: 216–221.
32. Graham SP, Honda Y, Pellé R, Mwangi DM, Glew EJ, et al. (2007) A novel strategy for the identification of antigens that are recognised by bovine MHC class I restricted cytotoxic T cells in a protozoan infection using reverse vaccinology. *Immunome Res* 3: 2.
33. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor Laboratory Press: Cold Spring Harbor. New York.
34. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16: 276–277.
35. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
36. Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16: 111–120.
37. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
38. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6: 288–295.
39. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9: 299–306.
40. Yang Z, Nielsen R, Hasegawa M (1998) Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol Biol Evol* 15: 1600–1611.
41. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586–1591.
42. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22: 1107–1118.
43. Egan A, Mahurkar A, Crabtree J, Badger JH, Carlton JM, Silva JC (2008) IDEA: Interactive Display for Evolutionary Analyses. *BMC Bioinformatics* 9: 524.
44. Kumhaek C, Phra-ek K, Renia L, Singhasivanon P, Looareesuwan S, et al. (2005) Are extensive T cell epitope polymorphisms in the *Plasmodium falciparum* circumsporozoite antigen A leading sporozoite vaccine candidate selected by immune pressure? *J Immunol* 175: 3935–3939.
45. McKeever DJ (2009) Bovine immunity - a driver for diversity in *Theileria* parasites? *Trends Parasitol* 25: 269–276.
46. Freeman AR, Hoggart CJ, Hanotte O, Bradley DG (2006) Assessing the relative ages of admixture in the bovine hybrid zones of Africa and the Near East using X chromosome haplotype mosaicism. *Genetics* 173: 1503–1510.
47. Radley DE, Young AS, Grootenhuys JG, Cunningham MP, Dolan TT, et al. (1979) Further studies on the immunization of cattle against *Theileria lawrencei* by infection and chemoprophylaxis. *Vet Parasitol* 5: 117–128.